

Gender Development in Transgender Preschool Children

Anne A. Fast, and Kristina R. Olson
University of Washington

An increasing number of transgender children—those who express a gender identity that is “opposite” their natal sex—are socially transitioning, or presenting as their gender identity in everyday life. This study asks whether these children differ from gender-typical peers on basic gender development tasks. Three- to 5-year-old socially transitioned transgender children ($n = 36$) did not differ from controls matched on age and expressed gender ($n = 36$), or siblings of transgender and gender nonconforming children ($n = 24$) on gender preference, behavior, and belief measures. However, transgender children were less likely than both control groups to believe that their gender at birth matches their current gender, whereas both transgender children and siblings were less likely than controls to believe that other people’s gender is stable.

Gender is perhaps *the* central way in which children and adults carve the social world into categories (Maccoby, 1998; Ruble, Martin, & Berenbaum, 2006). Therefore, it may be unsurprising that gender is likely the earliest identity and social category to emerge in development (Lewis & Brooks-Gunn, 1979), and that acquiring gender knowledge is considered a critical component of early childhood development (Ruble et al., 2007). A pervasive, albeit an often implicit, assumption in society and in psychological research is that one’s gender (one’s sense of identity as a boy or girl) aligns with one’s sex (determined by one’s anatomy and chromosomes at birth). This belief is clearly grounded in data—for most people, their gender identity aligns with their sex. However, it is not *always* the case; rather there are people, termed *transgender*, whose gender identity and sex at birth do not align. One example is reality star, Jazz Jennings, who expressed a female identity as soon as she could communicate that information to others despite being born a natal male (Goldberg & Adriano, 2007). When she was 5 years old, her parents allowed her to begin living as a girl in everyday life (meaning that they used the pronoun “she” and a new female name “Jazz,” but no medical or hormonal intervention occurred at that age)—a process

called a *social transition*. In the current study, we ask whether children like Jazz show patterns of gender development within the early preschool years that are similar to or different from gender-typical children of the same age.

Despite developmental psychology’s long and rich history of studying gender development, children like Jazz—socially transitioned transgender children—have largely been absent from these investigations. This is in part because social transitions early in development are relatively new (Ehrensaft, 2011; Hidalgo et al., 2013). However, the unique developmental experiences of transgender children, especially those who “switch” their gender presentations early in life, may contribute in interesting ways to discussions about how gender and sex function as organizing principles in young children’s lives. Although their experiences are rare (estimates of transgender identities are difficult to find, but one recent study of New Zealander high school students suggested a rate of approximately 1.2% of people identifying as transgender, Clark et al., 2014; and likely even fewer have socially transitioned to live as the “other” gender), given that socially transitioned transgender children do exist, it is important to include their experiences in the study of gender development. Thus, consistent with arguments concerning the importance of increasing diversity in empirical psychology (e.g., Kang & Bodenhausen, 2015; Shelton, 2000), the inclusion of transgender children will further our understanding of the range

This research was supported by NSF Grant 1523632 and the Arcus Foundation. Any opinions, findings, or conclusions expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Arcus Foundation. The authors wish to thank Madeleine DeMeules, Lily Durwood, Elizabeth Ake, Sarah Colombo, and Allison Skinner for assistance with data collection.

Correspondence concerning this article should be addressed to Anne A. Fast, Department of Psychology, University of Washington, Seattle, WA 98195. Electronic mail may be sent to afast@uw.edu.

© 2017 The Authors

Child Development © 2017 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2017/xxxx-xxxx

DOI: 10.1111/cdev.12758

of ways in which gender emerges and develops while also offering possible contributions to theoretical discussions of gender development (Dunham & Olson, in press). In the current study, we aim to do so by investigating preschool-age socially transitioned transgender children's gendered preferences, behaviors, and beliefs. We discuss how these data can add to our understanding of gender development, inform theories of gender development, and give rise to new research questions concerning the development of gender cognition.

Theoretical Contribution

Beyond conducting an exploratory analysis of the basic gender development of socially transitioned transgender children and comparing it to the development of gender-typical children of the same age, a secondary goal of the current study was to provide data that can begin to speak to broader theoretical discussions about gender development and transgender children. As one example, some previous studies has claimed that understanding gender constancy (that gender is a stable and consistent attribute) allows for greater organization and motivation of strong same-gender preferences and behaviors (Kohlberg, 1966; Slaby & Frey, 1975). More recent theorizing, however, suggests that although full gender constancy knowledge may not be responsible for enhancing gendered preferences and behavior, understanding gender *stability* in particular is a central factor in motivating strong gender preferences (Ruble et al., 2007). For example, a very young girl might already display an affinity toward pink, dolls, and dresses, but once she understands the stability of her gender, she will have even more extreme gendered preferences.

Socially transitioned transgender children present an interesting case to test this idea because unlike other children, they might *not* have a belief that their gender is stable. Anecdotally, many transgender children and their families discuss how they "used to be" one gender but are another gender now, after their social transition. Although this conversation can be interpreted as the child having been *assumed to be* one gender and now being recognized as a member of the "other" gender, children may not have this nuanced understanding. Furthermore, as discussed in more detail below, there is some initial evidence suggesting that gender nonconforming children (a group that would include transgender children) are less likely than gender-typical children to say that gender is stable over time (Zucker et al., 1999). At the same time,

some work with older socially transitioned transgender children suggests that they give gender-typical (but not sex-typical) responses on measures of gender development, such as gendered preferences (Olson, Key, & Eaton, 2015). Evidence that transgender children show strongly gendered preferences (perhaps as strong as controls) paired with a lack of gender stability beliefs (at least as it has traditionally been tested) could suggest that the "boost" from stability beliefs is not needed to show the high levels of gendered preferences observed by gender-typical children.

In addition, this study is likely to spawn further theoretical work on questions about gender development that may not come about until we know how socially transitioned transgender children respond compared to their gender-typical peers. For example, if transgender children do not differ from gender-typical children on some or all measures of gender development, we would have some preliminary evidence that gender of rearing in the first few years may not be a large contributor to those particular aspects of gender development—a hypothesis we may then be able to test with future data collection with transgender children or children with other diverse early experiences, such as intersex children who were reared as one gender but later identified as the "opposite" gender. Thus, we see this study as a catalyst for the establishment of future research and theories of gender development.

Gender-Typical Development

A large body of research suggests that gender-typical children, or those whose gender identity aligns with their sex at birth, are attuned to cues about gender early in development and begin perceiving gender categories at a young age. Infants display an ability to discriminate male and female faces by 6 months of age (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002), and they can accurately match male and female voices to male and female faces by their first birthday (Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994). Around age 2, when they begin to acquire knowledge of gender labels (Fenson et al., 1994; Stennes, Burch, Sen, & Bauer, 2005), infants display preferences for objects and people associated with their own gender (Serbin, Poulin-Dubois, Colburne, Sen, & Eichstedt, 2001; Zosuls et al., 2009) and show rudimentary gender stereotyping (Levy & Haaf, 1994; Serbin, Poulin-Dubois, & Eichstedt, 2002).

The current study focuses on the preschool period, as it is a crucial time for gender development.

Throughout these years (3–5 years of age), gender is highly salient and a powerful motivator of children's preferences and behaviors. For example, preschool-age children use gender to guide their own outfit choices (Halim et al., 2014) and toy choices (Eaton, Von Bargen, & Keats, 1981), such that they express interest in objects that are associated with their own gender rather than those linked to the other gender. Similarly, by age 3 and throughout the preschool years, children display a strong preference for same-gender people (Martin & Fabes, 2001; Martin, Fabes, Evans, & Wyman, 1999; Shutts, Pemberton, & Spelke, 2013). Preschool-age children also use gender to guide their expectations of others' appearances and activities (Miller, Lurye, Zosuls, & Ruble, 2009), with gender stereotype knowledge developing rapidly during this age range (see Signorella, Bigler, & Liben, 1993). Finally, the preschool years are also thought to be a critical time for developing knowledge of gender constancy (Kohlberg, 1966; Slaby & Frey, 1975). Specifically, children at this age are thought to master an understanding that gender is *stable* from infancy to adulthood (not until after preschool are they thought to understand that gender is consistent across changes in appearance, Ruble et al., 2007). Moreover, this gender stability knowledge is considered to be a central factor in enhancing preschool children's same-gender preferences and behavior (e.g., Ruble et al., 2007).

Gender Nonconforming Children

Despite the large body of work on gender development, most of that work has been conducted with gender-typical children. The current study is an exploratory investigation into whether socially transitioned transgender children show the same patterns of gender development during the preschool years. To date, there has been no work on this question, though there has been one study reporting on gender development in elementary school-age children and a few studies reporting on gender development in a broader range of gender diverse children.

Olson et al. (2015) investigated a similar question about gender development milestones in elementary school-age transgender children ($M_{\text{age}} = 9$ years, 1 month). They found that across several measures—preferences for same-gender peers and objects endorsed by those peers, as well as in the degree to which they displayed an implicit or explicit gender identity, and implicit gender-based preferences—socially transitioned transgender children

did not differ from gender-typical control children (who were matched on age and expressed gender) and gender-typical siblings when considered according to their gender. When analyzed as a function of gender assigned at birth (i.e., according to natal sex), the transgender children differed from their controls and siblings on every measure. Thus, from Olson et al. (2015) we can conclude that, by the elementary years, socially transitioned transgender children show gender-typical responding on many measures of gender development.

In addition to this more recent work on socially transitioned transgender children, there is a longer tradition of studying gender nonconforming children—those who defy cultural gender expectations for children of their sex—in the clinical psychology and psychiatry literature. Although typically focused on clinical outcomes (e.g., Cohen-Kettenis, Owen, Kaijser, Bradley, & Zucker, 2003), these research teams have occasionally reported on basic gender development, much as we do in the current study with socially transitioned transgender children. For example, one study found that while siblings of gender nonconforming children preferred to play with toys manufactured for children of their sex, gender nonconforming children ($M_{\text{age}} = 7.6$ years) did not—they equally preferred toys manufactured for their own sex and the other sex (Zucker, Bradley, Doering, & Lozinski, 1985). In contrast, for games, gender nonconforming children actually expressed a preference for games manufactured for the other sex, but their siblings did not. Thus, it appears that gender nonconforming children's preferences consistently differ from their gender conforming peers (the definition of gender nonconformity), but only sometimes was this difference in the direction opposite their sex at birth. Furthermore, Zucker et al. (1999) found that a group of 3- to 10-year-old gender nonconforming children showed an atypical understanding of gender constancy. That is, the gender diverse group of children was less likely than gender-typical children to believe that their own gender was stable across time (gender stability) or across changes in appearance (gender consistency) compared to others' gender.

Importantly, although these previous findings suggest that gender nonconforming children have response patterns differing from their same-sex peers, these studies were completed with children who differ from those in the current study in two key ways. First, the children in this past work (aside from Olson et al., 2015) were not socially transitioned. That is, these children presented in public (e.g., attended school) as the gender that aligns with

their sex at birth. In contrast, the current study focuses on children who have socially transitioned, and thus present a gender that differs from the one they were assumed to have at birth. Second, previous studies with gender nonconforming children likely drew a broader range of children than those included in the current study. That is, there are many gender nonconforming children who do not actually believe themselves to *be* the “other” gender; instead, they tend to have preferences that align with the “other” gender while maintaining that they identify as a member of the gender group aligning with their sex at birth (see Olson, 2016 for more on this issue). Children in the current study—those expressing that they *are* a member of the “other” gender group and living publically as this “other” gender—perhaps could be thought of as the most extreme subset of gender nonconforming youth. Therefore, it may be reasonable to assume that these socially transitioned transgender children are the most likely to show effects in the direction opposite their sex at birth, leading to our hypothesis that these children would show patterns of gender responding remarkably similar to children who share their expressed gender but who differ in their sex (as Olson et al., 2015 found with older children).

The Current Study

The present study is an exploratory investigation comparing the gender preferences, behavior and beliefs of young socially transitioned transgender children to those of their gender typical peers. To this end, we had socially transitioned transgender children and gender-typical children complete a series of basic gender development tasks, including measures assessing: (a) gender constancy understanding, (b) gender preferences, (c) gendered behavior (stereotypicality of outfit worn to the appointment), (d) explicit gender identity, (e) perceived similarity to boys and girls, and (f) gender stereotyping.

As described in the Theoretical Contribution section earlier, assessing how socially transitioned transgender children respond compared to gender-typical peers on measures of gender constancy understanding—more specifically, gender stability understanding—paired with their responding to measures of gender preferences may shed some light on whether seeing one’s gender as stable over time is always a major contributor to enhancing same-gender preferences and behavior at this age. Based on both anecdotal knowledge and previous work with gender nonconforming children (Zucker

et al., 1999), we expected that young transgender children may be less likely than gender-typical children to see their gender as stable over time. However, drawing upon previous work on older socially transitioned transgender children (Olson et al., 2015), we expected young transgender children to have just as strong same-gender preferences and behavior as their gender-typical peers. If transgender and gender-typical children show similar levels of preference but different patterns of stability, this might suggest that stability is not playing a causal role, so much as it is a developmentally co-occurring phenomenon.

Because previous work on gender development has included questions about gender identity, we do so in the current study as well; however, the measures we use are slightly different from those used in past work. The explicit gender identity measure used in the current study asks children what they feel like they are on the inside (i.e., in their “mind, thoughts, and feelings”). This identity measure was designed to make it clear to children that we were asking about their *gender identity* and not their biological sex. We also investigated gender identity by asking children how similar they feel to boys and how similar they feel to girls. Finally, we explored gender stereotyping (vs. flexibility). Because there is no previous work on transgender or gender nonconforming children’s tendency to endorse gender stereotypes, this was an exploratory investigation.

In the current study, we include two control groups, serving two different purposes. The first is a group of gender-typical children that are matched to the transgender children on age and gender identity (henceforth, *controls*). This group allows us to examine whether our transgender sample is showing responses typical for their age and gender. The second control group is a group of gender-typical children who are *siblings* of transgender or gender nonconforming children (henceforth, *siblings*). The inclusion of the sibling control group allows us to separate the impact of the lived experience of gender diversity from mere knowledge of the existence of gender diversity. In the case that transgender children respond differently from controls *and* siblings, we might assume personal experience of being transgender is playing a critical role in this difference. On the other hand, a finding that siblings respond similarly to transgender children but different from controls would suggest that knowledge of gender diversity (or perhaps factors unique to the kinds of families that have gender nonconforming children in them) plays a contributing role rather than just personal experience as a transgender person.

We specifically focused on socially transitioned transgender children in the preschool years (ages 3–5) because it is the age at which—according to previous work with gender-typical children (Eaton et al., 1981; Halim et al., 2014; Shutts et al., 2013; Signorella et al., 1993)—gender begins to strongly motivate children’s preferences and behavior as well as their beliefs about gender roles. Additionally, the preschool years are when an understanding of gender stability is mastered (Ruble et al., 2007; Slaby & Frey, 1975), which is particularly interesting with regard to a young transgender person’s view of their gender identity. In addition, this is the age at which gender nonconforming children tend to express their gender-atypical identity and behavior (Green, 1976; Zucker, Bradley, & Sanikhani, 1997), and it is the age at which the first transgender children have socially transitioned (to our knowledge).

transitioned transgender children, currently ages 3–14. This study focused on the 3- to 5-year-old children in that project. Participants in this study belong to three different groups: (a) socially transitioned binary (meaning they identify as male or female) transgender children (henceforth, *transgender*), (b) gender-typical siblings of transgender and gender nonconforming children, and (c) age- and gender-matched unrelated gender-typical control children. Because transgender children are rare, the research team traveled extensively to recruit this sample. Over the course of 9 months (March 2015 to November 2015), the researchers flew and drove throughout the United States to meet with families from 17 U.S. states (see Table 1 for list of states) at a series of conferences and camps for gender diverse children, at support group meetings for families with transgender children, at our research laboratory (for area families), or at families’ own homes to recruit this sample of transgender children and siblings. Additionally, control participants were run in a child development laboratory in the Pacific Northwestern United States. Despite recruiting from different geographic areas, our groups had similar other demographics as can be seen in Table 1.

Method

Participants

This study is part of a larger longitudinal project on gender development, examining the longitudinal development of a larger sample of socially

Table 1
Participant Family Demographics

	Controls	Siblings	Transgender	Statistic	<i>p</i> Value
Child’s age ^a	5.03 (8.52)	4.93 (8.24)	4.99 (7.82)	$F(2, 93) = 0.15$.858
Child’s gender ^b					
Male	22%	50%	22%	$\chi^2(2) = 6.72$.035
Female	78%	50%	78%		
Child’s race ^c					
White	78%	75%	61%	$\chi^2(2) = 1.87$.392
Monoracial, non-White	8%	4%	6%		
More than one	14%	21%	33%		
Household annual income ^d					
Less than \$25,000	3%	0%	3%	$F(2, 93) = 0.46$.632
\$25,001 to \$50,000	0%	17%	5%		
\$50,001 to \$75,000	19%	0%	8%		
\$75,001 to \$125,000	36%	50%	31%		
More than \$125,000	42%	33%	53%		
Parental political ideology ^e	2.35 (1.11)	1.54 (0.66)	1.64 (0.76)	$F(2, 93) = 8.07$.001
U.S. state of residence ^f	WA	AZ, CA, CO, IL, IN, MA, MD, MO, NJ, OH, OR, VT, WA, WI	AZ, CA, CO, DC, IL, IN, MA, MD, NM, OR, PA, WA		

Note ^aMean age (years) with standard deviation (months) of child participants. One-way analysis of variance was conducted on child’s age in months. ^bPercentage of male-identified and female-identified children. ^cPercentage of child in each racial category. Chi-square analysis was conducted with two categories (White and non-White). ^dPercentage of families in each income bracket. One-way analysis of variance was conducted on a 5-point scale variable that maps onto the five options provided. ^eMean and standard deviation of parents’ political ideology on a scale ranging from 1 (*very liberal*) to 7 (*very conservative*). ^fCurrent U.S. states of residence for participants in each group.

Transgender Children

Thirty-six 3- to 5-year-old transgender children who had socially transitioned ($M_{\text{age}} = 4.99$ years, $SD = 7.82$ months) participated, including 28 transgender girls (natal males) and 8 transgender boys (natal females). Perhaps not surprisingly, as social transitions often occur later in development, our sample skewed toward the older age of this range, with two 3-year-olds, thirteen 4-year-olds, and twenty-one 5-year-olds participating. The transgender children were socially transitioned at the time of participation, meaning they were all living as the gender “opposite” of their natal sex. Using the criteria for full transitions from Steensma, McGuire, Kreukels, Beekman, and Cohen-Kettenis (2013), participants had to be using the pronoun, clothing, and hairstyles associated with the “other” gender to count as socially transitioned. Every socially transitioned transgender 3- to 5-year-old child who was run during the recruitment period is reported in this article.

Siblings

In order to recruit as large of a sample of siblings as possible, we recruited a group of children who were siblings of transgender children, irrespective of whether the transgender sibling was in the 3- to 5-year-old range (only two of the current siblings had one in this sample), or who were siblings of children who were gender nonconforming (i.e., children who had behaviors and preferences counter to gender stereotypes and who had not yet transitioned). Twenty-four 3- to 5-year-old gender-typical siblings of transgender or gender nonconforming children ($M_{\text{age}} = 4.93$ years, $SD = 8.24$ months; 12 natal males, 12 natal females) participated during the same time period as the transgender children. These children were recruited and run while attending the same support group meetings, conferences, and camps as our transgender sample and were recruited via the same recruitment techniques, as we always stated that we were interested in sibling participants as well. All siblings age 3–5 years old who were run during the recruitment period are included in this article.

Controls

Additionally, thirty-six 3- to 5-year-old gender-typical children ($M_{\text{age}} = 5.03$ years, $SD = 8.52$ months; 8 natal males, 28 natal females) were recruited to participate as matched controls of the

transgender participants. These controls were matched on age, such that controls’ ages at test were within 4 months of the transgender children’s age at test, and matched on expressed gender (such that a transgender girl—a natal male who lives as a girl—was matched to a gender-typical girl), which is the same matching approach as utilized by Olson et al. (2015). Gender-typical matched controls were recruited through a university database of families interested in participating in child development research, and families were informed this was part of a study of children with diverse gender identities and expressions.

Procedure and Materials

After obtaining parental consent and verbal assent, experimenters administered the following measures. Parents simultaneously provided demographic and other information. The procedures took approximately 30 min to complete. All participants, including parents, received \$10, and children additionally received a small toy prize for participation.

Gender Constancy

Participants were asked questions about their own and others’ gender stability and consistency to examine their gender constancy understanding. We chose to separate first-party and third-party gender constancy because previous work on gender constancy understanding of gender-atypical children examined constancy knowledge in this way (Zucker et al., 1999). Furthermore, it was plausible that transgender children would view the constancy of their own gender differently compared to the gender constancy of others—after all, unlike children themselves, most people the children know have had a stable gender.

First-party stability. Two questions about participants’ own gender stability were taken from a previous work (Slaby & Frey, 1975). Participants were asked about their gender in the past (“When you were a little baby, were you a little boy or a little girl?”) and in the future (“When you grow up, will you be a dad or a mom?”). Because we were interested in the degree to which transgender children responded that their gender was stable from the past compared to their responses about the stability of their gender going into the future, we separately analyzed this measure by item. To examine whether transgender children differ from siblings and controls in their pattern of responding to the two first-party stability questions, we coded

whether participants responded to each question with their expressed gender or with the “opposite” gender (i.e., the gender “opposite” of natal sex for gender-typical controls and siblings, and the gender that aligns with natal sex for transgender participants). Participants who did not provide a response to an item were excluded from analyses of that item, leading to the exclusion of one control participant, one sibling, and two transgender participants on the past stability item as well as one control participant and four transgender participants on the future stability item.

Third-party stability. To measure participants’ understanding of others’ gender stability, four questions were adapted from previously validated measures (Ruble et al., 2007; Szkrybalo & Ruble, 1999). Participants were shown pictures of four different targets (a boy, a girl, a woman, and a man) one at a time and answered one question about each target. When participants were shown a boy or a girl, they were asked, “When this kid was a little baby, was this kid a boy or a girl?” and when participants were shown a man or a woman, they were asked, “When this grown-up was little, was this grown-up a boy or a girl?” Participants’ responses to the four third-party stability questions were coded as 1 if they responded with the gender-constant response (e.g., saying that a boy will be a man) and 0 if they gave any other answer (e.g., opposite gender, both genders). These four variables (Cronbach’s $\alpha = .90$) were summed to create a *third-party stability total* score (with a possible range from 0 to 4, see Table 3 for means). Participants had to respond to all four third-party stability items to be included in analyses of this measure, resulting in the exclusion of one control participant and two transgender participants.

First-party consistency. Two questions were taken from previous work (Slaby & Frey, 1975) to assess participants’ own gender consistency. Participants were asked two questions: “If you wore [opposite gender’s] clothes, would you be a boy or a girl?” and “If you played [opposite gender’s] games, would you be a boy or a girl?” Critically, when presenting the consistency questions, we asked about the gender “opposite” children’s expressed identity (e.g., a gender-typical or transgender girl was asked about boys’ clothes or games). In this way, all children were asked about clothing that would have been less common for them to wear in their current everyday life, as the measure was originally designed to function in that way. Participants’ responses to the two first-party consistency questions were coded as 1 if they responded with

the gender corresponding to their expressed gender and 0 if they gave any other answer (e.g., opposite gender, both genders). These two variables (Cronbach’s $\alpha = .79$) were then summed to create a *first-party consistency total* score for each participant (with a possible range from 0 to 2, see Table 3 for means). Participants had to respond to both first-party consistency items to be included in analyses of this measure, which resulted in the exclusion of one control participant and three transgender participants.

Additionally, for these consistency items, children were asked to provide a justification of why they gave each response. The field is split about whether or not to code justifications given by 3- to 5-year-old children—some previous researchers code justifications to consistency questions (Arthur, Bigler, & Ruble, 2009; Ruble et al., 2007), but many do not (Bussey & Bandura, 1984; Frey & Ruble, 1992; Lobel & Menashri, 1993; Marcus & Overton, 1978; Slaby & Frey, 1975; Warin, 2000). Because our participants tended to give nonsense, “I don’t know,” or no justifications (40% did so at least once during the first-party consistency task, leaving very few responses that could be coded), and because 3- to 5-year-olds in our study and in past work generally do not pass gender consistency measures anyway, we did not use these justification responses for first-party or third-party consistency measures; doing so would necessarily mean children would perform even worse on these items.

Third-party consistency. To measure participants’ understanding of others’ gender consistency, four questions were adapted from previously validated measures (Ruble et al., 2007; Szkrybalo & Ruble, 1999). Participants were shown four new targets (a boy, a girl, a woman, and a man) and were asked a question about each target. When participants saw a boy or a girl, they were asked, “If this kid wore [opposite gender’s] clothes, would this kid be a boy or a girl?” and when participants saw a man or a woman, they were asked, “If this grown-up did the work that [opposite gender] do, would this grown-up be a man or a woman?” Participants’ responses to the four third-party consistency questions were coded as 1 if they responded with the gender-constant response (e.g., saying that a boy will be a man) and 0 if they gave any other answer (e.g., opposite gender, both genders). Then, these variables (Cronbach’s $\alpha = .87$) were summed to create a *third-party consistency total* score (with a possible range of 0–4, see Table 3 for means). Participants were required to respond to all four third-party consistency items in order to be included in

analyses of this measure, which resulted in the exclusion of two control participants, two sibling participants, and two transgender participants.

Preferences

Participants were asked about their peer, toy, and clothing preferences.

Peer preference. Participants saw eight separate pairs of children and were asked to point to the child they would like to be friends with the most (Olson et al., 2015). In six of the trials, the pair included a male child and a female child, matched on perceived age and attractiveness, while two filler trials included two apparently male children or two apparently female children. A *peer preference score* was calculated for each participant, representing the number of times on mixed-gender pair trials (0–6) the participants chose peers who were the gender that matched their own expressed gender (e.g., number of times a gender-typical girl or transgender girl picked girls). Participants who did not provide a response on every trial of the task (with the exception of the two filler trials) were excluded from analyses of this item, which resulted in the exclusion of one control (missed five of six items) and one transgender participant (missed four of six items).

Toy and clothing preferences. Participants saw four sets of five toys and four sets of five outfits and were asked to point to the toy they would like to play with the most or which outfit they liked the best. For example, one group of toy items included an orange tool set, a red barbecue set, a board game, a purple stove set, and a pink kitchen set. An example set of clothing items included a pair of plaid cargo shorts with athletic T-shirt, a pair of gray jeans with blue button-down shirt, a pair of blue jeans with green T-shirt, a pair of blue jeans with pink tank-top, and a purple dress with sparkles. Thus, each set of five toys or outfits could be arranged from 1 (*most stereotypically masculine*) to 5 (*most stereotypically feminine*). These items had previously been pilot tested with a group of gender-typical children to determine how stereotypically girl-like or boy-like they were, and in the current study, the items were found to be highly reliable (toy items: Cronbach's $\alpha = .74$; clothing items: Cronbach's $\alpha = .92$). Responses were averaged to create a *toy preference score* and a *clothing preference score*. For boys, scores were recoded such that higher numbers represent more gender-consistent preferences (the scale was already ordered that way for girls). Analyses included participants who responded to at least

three of the four items in each category, resulting in the inclusion of all participants (three children skipped one toy item—one control, one sibling, one transgender—and two transgender children skipped one clothing item, but due to the averaging approach, these participants could nonetheless be included in analyses).

For analysis purposes, these three measures were combined to create a *preferences composite score*. Because the measures were on two different scales, the peer preference (ranging from 0 to 6), toy preference (1 to 5), and clothing preference (1 to 5) scores were first standardized into percent of maximum possibility (POMP) scores (Cohen, Cohen, Aiken, & West, 1999). The POMP scores were calculated by first subtracting the minimum possible score on the scale from the observed scores. That difference was then divided by the difference between the maximum and minimum possible scores on the scale, which was then multiplied by 100. Once POMP scores were calculated for each preference score (peer, toy, and clothing), they were averaged to create the *preferences composite score*. Although we do report the means on the original scales in the table below, the analyses are conducted with the *preferences composite score*.

Stereotype Flexibility

A task was adapted from Liben and Bigler (2002) to assess the degree to which participants endorsed flexibility about gender activity stereotypes. Participants were told that they would hear a list of activities that people can do (e.g., gymnastics and video games; see Supporting Information for full list), and to say who they think *should* do each activity: boys, girls, or both boys and girls. Responses were coded into a *stereotype flexibility score*, which is the number of times each participant responded that “both boys and girls” should do an activity that was previously deemed either stereotypically male or stereotypically female. Because 5 of the 15 items were intended to be gender-neutral activities, those items were excluded from analyses. Thus, only the 10 items about gendered activities were included in participants' stereotype flexibility scores, which represent the number of times participants responded with the “both” option on gendered items (ranging from 0 to 10). Some participants did not even begin the measure (one control participant, two siblings, and two transgender participants), and thus are excluded from analyses on this measure. Of the participants who did start this stereotype measure, five did not respond to all of the questions,

resulting in the additional exclusion of four transgender participants and one sibling on this measure.

Explicit Gender Identity

Participants reported their gender identities using the explicit gender identity measure that Olson et al. (2015) used with elementary school-aged transgender children. Before answering gender identity questions, participants were told that everybody has an outside part (physical body) and an inside part (mind, thought, and feelings) of them. Participants were further told that for some people the outside and inside parts are the same, and for other people they are different. For example, a person could be a boy on the outside and feel like a boy on the inside or could be a boy on the outside and feel like a girl on the inside. Additionally, participants were told that some people feel like they are both, neither, or that it changes over time. Finally, participants reported (a) what they feel like on the *inside* right now, and (b) what they think they will feel like on the *inside* when they grow up: a boy or man, a girl or woman, neither, both, it changes over time, or I don't know. Participants who did not provide a response to either explicit identity item were excluded from analyses of that particular item (first (now) item: two transgender participants and two siblings; second (grown up) item: one control participant, two transgender participants, and two siblings).

Similarity

Participants completed a task developed by Martin, Andrews, England, Zosuls, and Ruble (2016), measuring how similar children think they are to boys and girls. Participants answered 10 questions (5 for similarity to boys and 5 for similarity to girls) and responded on a scale ranging from 0 (*very different*) to 4 (*very similar*). More specifically, participants were asked, "How similar do you feel to boys[girls]?" "How much do you act like boys[girls]?" "How much do you look like boys[girls]?" "How much do you like to do the same thing as boys[girls]?" and "How much do you like to spend time with boys[girls]?" The response scale included a visual representation of each option to help participants understand the possible responses. This visual representation displayed circles labeled "You" and "Boys"["Girls"], with the circles varying in the degree of overlap or separation, mapping on to the degree of (dis)similarity to other kids.

We calculated three scores for this similarity measure. First, we created a *similarity to my gender* score, which is an average of the five items about other kids with the same gender as the participant (Cronbach's $\alpha = .70$). For example, for transgender girls and gender-typical girls, the similarity to my gender score is the average of the items asking about similarity to other girls. Next, we calculated a *similarity to other gender* score, which is an average of the five about other kids with the "opposite" gender as the participant (Cronbach's $\alpha = .71$). For transgender girls and gender-typical girls, the similarity to other gender score would be the average of the items asking about similarity to boys, for example. We also created a *similarity difference* score by calculating the difference between the similarity to my gender score and the similarity to other gender score. This difference score was always calculated in the direction of the participant's expressed gender (e.g., scores for transgender and gender-typical girls were calculated by subtracting the "similarity to boys" from the "similarity to girls" average).

Participants had to answer all questions included in each composite score to be included in analyses of that score. Because this similarity measure was always the last task in the procedure, a number of participants did not even begin the measure (3 control participants, 7 siblings, and 10 transgender participants), and thus are excluded from analyses on this measure. Of the participants who did start this similarity measure, three missed at least one question contributing to each composite score, resulting in the additional exclusion of one control participant, one sibling, and one transgender participant from analyses of all scores for this measure. Finally, one additional transgender participant was excluded from analyses of only the *similarity to other gender* and *similarity difference* scores for not answering one of the questions about similarity to children of the other gender.

Outfit at Appointment

To measure participants' gender expression in everyday life, without telling parents or children in advance, two experimenters independently rated the outfit worn by each participant at the testing session on a scale ranging from 1 to 5 (allowing for half-point ratings) with lower numbers representing more stereotypical boy outfits and higher numbers representing more stereotypical girl outfits ($r = .94$, $p < .001$). However, in some cases ($n = 13$) only one experimenter was able to provide an outfit rating,

and in those cases we just used the one experimenter’s rating (unfortunately, due to experimenter error, for three participants—one transgender participant and two siblings—the experimenter did not indicate a rating, thus those three were excluded from analyses for this measure). Experimenters were told that the most masculine outfits consisted of clothing items such as male-stereotypic sports attire, superhero costumes, and men’s formal wear, whereas the most feminine outfits consisted of frilly dresses or skirts, princess costumes, and sparkly accessories. Experimenters also considered the colors (e.g., pink) and style (e.g., fitted vs. baggy shirt) when determining outfit ratings.

Other Measures

Importantly, and in the spirit of transparency (Simmons, Nelson, & Simonsohn, 2011), we note that these measures were given as part of a larger study about gender development and mental health among gender diverse children. Therefore, during this time period we did collect data on two measures not reported here. First, 72% of the children in this study completed a measure of gender essentialism. However, that measure was intended for a article in progress on essentialism, which also includes participants within a larger age range (i.e., children who are older and are not included in the current article). Second, we added a new measure—on gender encoding—part way through this study; however, this measure was only completed by 18% of our participants and as such, will be reported in a separate article. Furthermore, while the current participants were completing these measures, their parents completed a variety of measures (e.g., mental health), but as those measures were not relevant to the present article (which is focused on children’s own behaviors, beliefs, and attitudes), they have been excluded from this article as well.

Results

Effect sizes for analyses of variance (ANOVAs) and post hoc comparisons were calculated using a spreadsheet created by Lakens (2013).

Gender Constancy

First-Party Stability

Chi-square analyses on responses to the first-party stability questions indicated that the

participants in the three groups differed significantly in their tendency to say their expressed gender in response to the question about their *past* gender, $\chi^2(2) = 57.32, p < .001, \phi = .789$; however, participants in the three groups were no different in their tendency to say their expressed gender in response to the question about their *future* gender, $\chi^2(2) = .081, p = .960, \phi = .030$. In response to the question about their gender as a baby, only 21% of transgender participants said their expressed gender, whereas 97% of controls and 96% of siblings said their expressed gender. On the other hand, when asked about their gender as an adult, 97% of transgender participants, 97% of controls, and 96% of siblings replied with their expressed gender (see Table 2). To best understand this result, imagine a child like Jazz from the Introduction—a natal boy who identifies as a girl. If she was the modal participant in our study, she would have said she was a boy as a baby, but will be a woman as a grown-up.

First-Party Consistency

A one-way ANOVA on *first-party consistency total* scores indicated that participants in the three groups did not differ in the degree to which they believed their gender would remain consistent across situational changes, $F(2, 89) = 0.96, p = .389, \eta_p^2 = 0.02$ (see Table 3 for proportion of participants in each group who gave consistent responses). Within each group, children’s responses did not differ from chance responding: transgender, $t(32) = .96, p = .344, d = 0.17$; siblings, $t(23) = .46, p = .647, d = 0.09$; controls, $t(34) = .90, p = .377, d = 0.15$. Thus, consistent with past research examining preschool-age children, irrespective of whether they were transgender or not, children did not systematically believe gender was consistent across changes in appearance.

Table 2
Proportion of Participants Responding to First-Party Stability Items With Expressed Gender

	Proportion	N	$\chi^2(2)$	p Value
First-party stability past				
Controls	.97	35	57.32	< .001
Siblings	.96	23		
Transgender	.21	34		
First-party stability future				
Controls	.97	35	0.81	.960
Siblings	.96	24		
Transgender	.97	32		

Table 3
Means and Proportion of Participants "Passing" Constancy Items

	M	SD	N	Statistic	p Value
Third-party stability					
Controls	3.89	0.47	35	$F(2, 90) = 5.36$.006
Siblings	3.17	1.49	24		
Transgender	2.94	1.56	34		
First-party consistency					
Controls	0.86	0.94	35	$F(2, 89) = 0.96$.389
Siblings	1.08	0.88	24		
Transgender	1.15	0.91	33		
Third-party consistency					
Controls	1.76	1.81	34	$F(2, 87) = 2.69$.073
Siblings	2.68	1.56	22		
Transgender	2.53	1.54	34		

Note Third-party stability total scores and third-party consistency total scores have a possible range of 0–4. Possible range for first-party consistency total scores is 0–2.

Third-Party Stability

A one-way ANOVA on *third-party stability total scores* revealed that the three groups differed in the degree to which they believed other people's gender would remain stable over time, $F(2, 90) = 5.36$, $p = .006$, $\eta_p^2 = 0.11$. Tukey's honestly significant different tests indicate that transgender participants were significantly less likely to say that others' gender is stable over time compared to control participants, $p = .006$, $d = 0.83$, but they were not different from siblings, $p = .775$, $d = 0.15$. Siblings were marginally different from control participants in the degree to which they endorse gender stability in others, $p = .079$, $d = 0.71$ (see Table 3 for the proportion of participants in each group who gave stable responses). Overall, all groups were significantly more likely than chance to believe that gender would be stable: transgender, $t(33) = 3.53$, $p = .001$, $d = 0.60$; siblings, $t(23) = 3.83$, $p = .001$, $d = 0.78$; controls, $t(34) = 23.69$, $p < .001$, $d = 4.0$. Thus, while all groups generally believed that other people's gender was typically stable across time, transgender children and to a lesser extent, the siblings of transgender and gender nonconforming children, responded that occasionally another child's gender could change across their life span.

Third-Party Consistency

A one-way ANOVA on *third-party consistency total scores* revealed a marginal difference between groups in participants' tendency to say that others' gender is consistent across situational changes,

$F(2, 87) = 2.69$, $p = .073$, $\eta_p^2 = 0.06$. However, if anything, the mean scores were higher (indicating greater belief in consistency) among the sibling and transgender groups compared to the control group (see Table 3 for proportion of participants in each group who gave consistency-relevant responses). Responses from control participants did not differ from chance responding, $t(33) = .76$, $p = .454$, $d = 0.13$; however, transgender participants were marginally more likely than chance to say that others' gender is consistent, $t(33) = 2.00$, $p = .054$, $d = 0.38$, as well as siblings, $t(21) = 2.06$, $p = .052$, $d = 0.44$. In sum, transgender children and their siblings, but not control participants, trended toward believing that another person's gender was likely to be consistent across changes in appearance.

Preferences

We conducted a one-way ANOVA on the *preferences composite score* to test whether transgender, sibling, and control participants differed in the degree to which they prefer same-gender peers, toys, and clothing (see Table 4 for means on original scales and *preferences composite*). Participants in the three groups did not differ in the degree to which they prefer same-gender peers and items, $F(2, 91) = 2.21$, $p = .116$, $\eta_p^2 = 0.05$. In all groups, children were significantly more likely than chance to prefer peers and items in the direction of their own gender: transgender, $t(34) = 15.53$, $p < .001$, $d = 2.62$; siblings, $t(23) = 4.68$, $p < .001$, $d = 0.96$; controls, $t(34) = 10.43$, $p < .001$, $d = 1.76$.

Stereotyping

A one-way ANOVA on *gender stereotype flexibility scores* (ranging from 0 to 10, see Table 4 for means) suggests that the transgender, sibling, and control participant groups did not differ on how much they endorse flexibility of gender stereotypes, $F(2, 83) = 0.77$, $p = .464$, $\eta_p^2 = 0.02$. In all groups, participants were significantly more likely than chance (3.33) to say that both boys and girls should do the gendered activities: transgender, $t(29) = 5.52$, $p < .001$, $d = 1.0$; siblings, $t(20) = 3.86$, $p = .001$, $d = 0.84$; controls, $t(34) = 4.75$, $p < .001$, $d = 0.80$. Figure 1 displays responses broken down by participant group and type of stereotype item.

Gender Identity

In terms of gender identity, we categorically coded participants' responses, such that participants

Table 4
Preference, Stereotyping, Similarity, and Outfit Means by Group

	M	SD	N
Peer preference ^a			
Controls	4.94	1.63	35
Siblings	4.29	1.90	24
Transgender	5.09	1.36	35
Toy preference ^b			
Controls	3.96	0.66	36
Siblings	3.67	0.94	24
Transgender	3.55	0.85	36
Clothing preference ^b			
Controls	4.49	0.88	36
Siblings	4.14	1.10	24
Transgender	4.65	0.59	36
Stereotype flexibility ^c			
Controls	5.54	2.76	35
Siblings	6.29	3.51	21
Transgender	6.43	3.08	30
Similarity to my gender ^d			
Controls	3.08	0.87	32
Siblings	3.10	0.87	16
Transgender	2.97	0.94	25
Similarity to other gender ^d			
Controls	1.13	0.96	32
Siblings	1.20	1.11	16
Transgender	0.99	0.81	24
Similarity difference ^e			
Controls	1.95	1.37	32
Siblings	1.90	1.33	16
Transgender	2.02	1.50	24
Outfit at appointment ^f			
Controls	4.14	0.59	36
Siblings	4.13	0.66	22
Transgender	4.25	0.63	35

Note ^aPeer preference scores range from 0 to 6, with higher numbers representing greater preference other children of one's own gender. ^bToy and clothing preference scores range from 1 to 5, with higher numbers representing more gender-typed preferences. ^cStereotype flexibility scores range from 0 to 10, indicating the number of times children said "both" in response to the stereotypic activities, such that higher numbers representing greater flexibility. ^dSimilarity to my gender and similarity to other gender scores range from 1 to 5, representing the degree to which participants think they are similar to children of the same gender as them or the other gender, respectively. ^eSimilarity difference scores represent the degree to which participants think they are similar to children of their own gender compared to children of the other gender (similarity to my gender–similarity to other gender). ^fOutfit at appointment scores range from 1 to 5, with higher numbers representing more gender-typed outfits.

could either respond with (a) their expressed gender, (b) the "opposite" of their expressed gender (opposite of natal sex for controls and siblings and same as natal sex for transgender participants), or (c) one of the additional options (i.e., neither, both,

it changes over time, or I don't know). Chi-square analyses revealed that transgender, sibling, and control participants did not differ in their likelihood of responding with their expressed identity when asked about both their current gender identity, $\chi^2(4) = 3.19, p = .526, \phi = .186$, and future gender identity, $\chi^2(4) = 3.77, p = .438, \phi = .204$ (see Table 5 for a summary of participant responses to both the current and future gender identity questions).

Similarity

The transgender, sibling, and control groups also not differ on how similar they felt to children of their same gender, $F(2, 70) = 0.15, p = .861, \eta_p^2 = 0.004$, one-way ANOVA, and how similar they felt to children of the other gender, $F(2, 69) = 0.26, p = .768, \eta_p^2 = 0.007$, one-way ANOVA (see Table 4 for means). The three groups also did not differ on how similar they feel to their own gender versus the other gender, $F(2, 69) = 0.04, p = .966, \eta_p^2 = 0.001$, one-way ANOVA. All groups tended to see themselves as similar to their own gender more than to the other gender: transgender, $t(23) = 6.60, p < .001, d = 1.35$; siblings, $t(15) = 5.70, p < .001, d = 1.43$; controls, $t(31) = 8.06, p < .001, d = 1.42$.

Outfit at Appointment

A one-way ANOVA on outfit ratings (ranging from 1 to 5, see Table 4 for means) indicated that transgender, sibling, and control participants did not differ in the degree to which they wore outfits stereotypically consistent with their expressed gender to the appointment, $F(2, 90) = 0.39, p = .680, \eta_p^2 = 0.009$. Children from all groups, on average, wore clothing that is stereotypically associated with their gender: transgender, $t(34) = 11.74, p < .001, d = 1.98$; siblings, $t(21) = 8.02, p < .001, d = 1.71$; controls, $t(35) = 11.64, p < .001, d = 1.94$.

Discussion

Across all measures of preference, behavior, stereotyping, and identity, if coded according to children's expressed gender, preschool-age socially transitioned transgender children never significantly differed from their gender-matched peers (age- and gender-matched controls and preschool-age siblings of transgender or gender nonconforming children), mirroring a previous finding about preferences and identity with older socially transitioned transgender

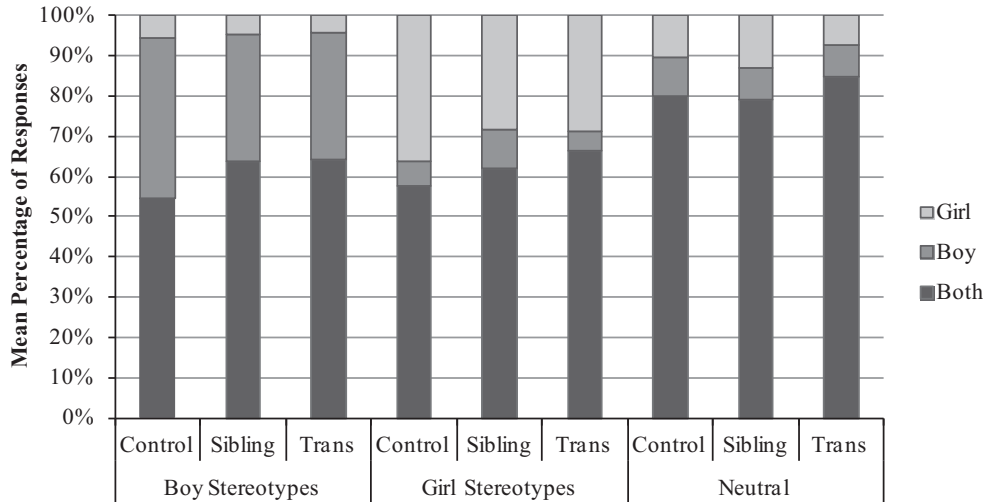


Figure 1. Mean percentage of responses to stereotype items by participant group and item type.

Table 5
Explicit Gender Identity Response Percentages by Group

	Controls (%)	Siblings (%)	Transgender (%)
Current gender identity			
Expressed gender	67	55	73
“Opposite” of expressed gender	8	5	6
Anything else	25	40	21
Future gender identity			
Expressed gender	66	50	68
“Opposite” of expressed gender	6	5	0
Anything else	28	45	32

Note The “anything else” response category includes the responses “neither,” “both,” “it changes over time,” and “I don’t know.”

children (Olson et al., 2015). That is, young transgender children were just as likely as gender-typical children to (a) show preferences for peers, toys, and clothing culturally associated with their expressed gender, (b) dress in a stereotypically gendered outfit, (c) endorse flexibility in gender stereotypes, and (d) say they are more similar to children of their gender than to children of the other gender. Transgender children were also just as likely as controls and siblings to say that they identify with their expressed gender, both now and in the future, when given multiple other choices. These findings suggest that, in many ways, the basic gender development of socially transgender children is quite similar to that of other children.

However, in terms of children’s responses to the gender constancy measures, the results were fairly mixed. Transgender children differed from the other children in that they tended to say that they were a different gender as an infant than their current gender in everyday life. However, they were just as likely as both control groups to say that their gender in adulthood would be congruent with their current gender. This pattern of responding to the gender stability items may reflect the way that transgender children’s families often talk about their gender—that everyone believed them to be one gender (the one associated with their sex), but now and in the future they have a different gender. Whether the family discussion reflects, leads to, or merely co-occurs with this responding is currently unclear; however, this pattern of responding is notably different from the way many binary-identified (male or female) transgender adults often discuss their gender—as always existing in this one way. One possible explanation for this difference is that children are interpreting the statement about their identity as an infant as being a question about sex, rather than gender, or that they are responding by considering how other people treated them rather than how they felt. If older transgender children, particularly those approaching puberty, are better able to interpret these gender stability items due to stronger awareness of the distinction between sex and gender or better separating their own beliefs about their gender from the beliefs others had about their gender, perhaps they would be more likely than younger transgender children to say that their gender has been stable across their

whole life. Future work can investigate this question, as well as assess whether preschool-age transgender children change the way they think about their gender as a young child based on how the question is presented (i.e., framed in terms of gender or sex). Transgender children did not differ from the control groups in thinking about the consistency of their gender identity across superficial changes (e.g., clothing), though none of the groups showed anywhere near ceiling level performance on these items.

Interestingly, transgender children were less likely to see *other people's* gender as stable over time compared to gender-typical controls. Although this finding could at first be seen as support for the claim that transgender children have quite a different understanding of gender than their gender-typical peers, the fact that transgender children did not differ from siblings on their third-party stability responding suggests instead that this effect may be the result of knowledge that gender is not stable over time for some people. Importantly, children in all three groups generally believed that gender would be stable across development for most people, meaning that even the transgender children and siblings made this assumption. The difference was that the transgender and sibling groups seemed to assume that occasionally there is an individual for whom this is not the case. Future research could examine whether knowledge of transgender people is causally related to this pattern of responding by teaching a group of gender-typical children about transgender children and then later assessing their third-party gender stability. With regard to third-party consistency, we found that transgender children were actually more likely than the control participants to respond that other people's gender was consistent in identity across clothing and hairstyle changes; though the difference was not quite significant, transgender children differed from chance responding, while controls did not. This finding could reflect transgender children's knowledge of the fact that gender identity can exist irrespective of what one wears since the children personally experienced a time during which they wore clothes of a gender that did not match their gender identity. However, given the fact that this difference between groups was not significant and that the sample size was small, we hastate to draw a particularly strong conclusion on this point.

Taken together, the current work suggests that preschool-age transgender children display similar patterns of gendered responding in terms of their

behaviors, preferences, stereotypes, and real-life clothing choices to that of gender-typical children. The largest difference in responses from the participant groups was in the domain of constancy, where transgender children were less likely than the gender-typical groups to believe that their own gender will remain stable from infancy to adulthood. The fact that transgender children had atypical responses on gender constancy measures, but typical responding on measures of gender preference, behavior, and stereotyping, sheds some doubt on Kohlberg's (1966) claim that full gender constancy understanding enhances gendered beliefs, preferences, and behaviors, as well as more recent claims that gender *stability* understanding in particular can boost same-gender preferences and behavior gender typing (Halim et al., 2014; Ruble et al., 2007). Although the development of gender stability understanding seems to co-occur with increases in gender typing for gender-typical children, the current work suggests that this knowledge is not necessary, nor does it appear to be a central contributor to strong same-gender preferences and behavior (for more discussion and evidence on this point, see the associated Supporting Information). Of course the best test of this causal question would be to conduct a longitudinal study of even younger children, a test we hope researchers will conduct in the future.

Kohlberg (1966) and more recent theorists (Halim et al., 2014; Ruble et al., 2007) did not distinguish between past and future identity when discussing children's knowledge of gender stability. However, nearly all children in our transgender sample believed their current expressed gender will be their gender in the future, responded with gendered preferences, endorsed gender stereotypes, and chose gendered clothing to the same degree as gender-typical children. Therefore, perhaps it is reasonable to consider an amendment to this cognitive theory of gender development, such that children's sense of gender stability from the current moment into the future (with the removed assumption that gender is necessarily consistent with a child's sex at birth) is what motivates or contributes to especially strong gendered preferences and behavior, rather than the belief that past gender must be stable. Again, a longitudinal study would best answer the causal component of this question. In addition, given the current debate about the degree to which gender identity *is* stable in transgender children (Olson, 2016; Soh, 2016; Steensma, Biemond, de Boer, & Cohen-Kettenis, 2011; Steensma et al., 2013; Vilain & Bailey, 2015; Zucker & Bradley, 1995), it

would be interesting to connect children's reasoning about gender stability and their actual later life identity.

Limitations and Future Work

As with all work, but especially exploratory work like the present project, there are considerable limitations in interpretation, and suggestions for improvements to make in further work. First, with regard to our measures, we have some concerns with the gender identity measure, developed by Olson et al. (2015), for use with preschool-age children. Rather than utilize traditional questions about gender identity (e.g., are you a boy or a girl?), we opted to provide children with a range of possible responses (both, neither, it changes over time, I don't know) that do not conform to a binary view of gender identity. Unexpectedly, a very large number of siblings and controls responded with one of the nonbinary options. One interpretation of this result is that children have more diverse identities than we typically assume them to have. We are skeptical of this interpretation because (a) we doubt that such a large number of children, especially in the control groups, have nonbinary identities given that older children do not express these identities as much (Olson et al., 2015), and (b) anecdotally, sometimes the very same children who gave nonbinary answers on these items clearly identified themselves as a member of a binary gender group (always the one they lived in during everyday life) at other times in the visit (something we unfortunately did not regularly assess). Furthermore, these findings are at odds with the results of the similarity measure which showed that children generally identified with their gender group. Thus, another possibility is that these younger children—who could not read the response options—forgot the options that were listed first (boy, girl) and responded with the options that were listed later (the nonbinary ones) or simply liked the idea of providing a nontraditional answer. Perhaps future work could focus on just three options (e.g., boy, girl, something else) to allow for nonbinary selections, but maintain a lower number of options to reduce demand. Finally, this gender identity measure was designed to inquire about children's gender identity rather than their sex. Thus, the items were worded such that children were asked about what they are on the "inside" following an experimenter's statement that the "inside" is our mind, thoughts, and feelings, as opposed to the "outside," which is our body. This information about

"insides" and "outsides" was intended to make it clear that we wanted to know how they felt on the inside (in their mind), however, some may have interpreted this as a question about their body (inside their clothes), interpreting this question differently than intended.

Relatedly, we may have also encountered an issue with interpretation in our gender constancy measure. Unlike the gender identity measure, we did not attempt to specify whether we were referring to gender and sex when asking about past versus future identity (i.e., stability) and identity upon situational changes (i.e., consistency). Thus, our results on these measures could reflect the fact that young children may not understand sex and gender as distinct, and further, that we did not provide adequate explanation for this distinction. For example, the difference between transgender and gender-typical children's responses on the gender stability measure might have occurred because in order to answer "correctly," transgender children would have to understand the distinction between sex and gender, but gender-typical children would not. That said, it is also worth noting that some transgender children at this age experience considerable body dysphoria. This experience might suggest that young transgender children actually have a representation of the distinction between sex and gender (i.e., they know that more female-identified people do not have penises, hence their dislike of their penises). Furthermore, these findings may suggest that transgender children, and to a lesser extent their siblings, have a deeper or more complex understanding of sex and gender than other children. A more systematic examination of transgender children's representation of this distinction should be conducted in future work in order to begin to disentangle these issues.

These findings open up questions regarding not only how transgender children interpret the items for these particular measures or how they represent sex versus gender, but also more broadly whether children at this age understand the distinction between gender and sex. Although there have been some research programs aimed at addressing these questions (Bem, 1989; Volbert, 2000), future work might capitalize on the existence of transgender children to validate these measures, for example.

Despite these limitations to some of the measures utilized in the present study, in general, most measures appeared to be interpretable and consistent with past research for control participants, giving us greater confidence in their interpretation. The considerable similarity across our three samples on

many of these basic gender development measures also suggests that the earliest years of rearing—the time during which transgender children were raised as a gender “opposite” that of the one they currently live as—may have minimal impact on many of our measures, such as preferences, stereotypes, and behavior. Instead, perhaps these constructs develop for the first time during the preschool years, a time at which the transgender children in this study have already socially transitioned. Examining children who transition later would be incredibly interesting. Alternatively, perhaps these constructs develop even earlier and are based on some of the same underlying mechanisms that lead transgender children to identify as the “opposite” gender in the first place. Future empirical and conceptual work in this area is needed.

A second limitation of the current work is that, although the gender- and age-matched controls were not different from the transgender children on many participant and family demographic measures, the groups did differ on parent political ideology. More specifically, the parents of the transgender and sibling children reported being more politically liberal than the (also very liberal) parents of the control children. On one hand, it is possible that more liberal parents are more likely to (a) have transgender or gender nonconforming children, (b) allow their children to socially transition, or (c) sign up to participate in research. Alternatively, having a transgender or gender nonconforming child may make a parent more liberal (in fact, we have had parents anecdotally report this while filling out the political orientation item). However, despite differences between our groups on parental political orientation, we found remarkably few differences between groups on our other measures, suggesting that while political orientation could influence social transitioning or identifying a child as transgender, it is unlikely (at least within the liberal range) to hugely influence the measures assessed here. Our participant groups differed significantly on one additional demographic variable—the gender breakdown of each group. Critically, our control participants were matched to transgender participants based on gender. For example, a transgender girl (natal male) was matched with a gender-typical girl. Thus, importantly, there was no difference in gender of children in our control and transgender participant groups, which were made up of 80% girls and 20% boys. However, the sibling group was 48% girls and 52% boys. Importantly, because our variables were coded with respect to each child’s gender (i.e., higher numbers

representing more stereotypic responses for child’s gender) and because the sibling group never differed from both of the other groups, it is unlikely that the gender breakdown of groups influenced responding on the measures.

A third, more general limitation of the current work is that the children included in this study are quite unique, and therefore, generalizing from these results must be met with caution. Within the greater population, very few children are transgender, fewer socially transition (by the preschool years), and fewer still sign up for research studies. Therefore, it remains an open question how widely these results will generalize. Moreover, the small sample size of the current work is another general concern that limits our ability to observe significant differences between groups. On the other hand, although most measures of basic gender development revealed no differences between transgender and other children, considering the number of tests conducted, it is important to take caution when interpreting the significant effects. Conducting many statistical tests increased the possibility of Type I errors, making replication especially important.

Conclusions

As a society and as psychological researchers, we are increasingly aware that there are individuals who identify, early in development, as a gender other than the one aligned with their sex at birth. Therefore, such children should be included in work on basic gender development, as in the current work, in order to both expand our knowledge of gender developmental experiences and strengthen our (widely debated) theories of gender development. In that way, as psychologists, we can lead not only the field, but also broader societal culture, in understanding that gender and sex identities are much broader than have traditionally been studied.

References

- Arthur, A. E., Bigler, R. S., & Ruble, D. N. (2009). An experimental test of the effects of gender constancy on sex typing. *Journal of Experimental Child Psychology, 104*, 427–446. doi:10.1016/j.jecp.2009.08.002
- Bem, S. (1989). Genital knowledge and gender constancy in preschool children. *Child Development, 60*, 649–662. doi:10.2307/1130730
- Bussey, K., & Bandura, A. (1984). Influence of gender constancy and social power on sex-linked modeling.

- Journal of Personality and Social Psychology*, 47, 1292–1302. doi:10.1037/0022-3514.47.6.1292
- Clark, T. C., Lucassen, M. F., Bullen, P., Denny, S. J., Fleming, T. M., Robinson, E. M., & Rossen, F. V. (2014). The health and well-being of transgender high school students: Results from the New Zealand adolescent health survey (Youth'12). *Journal of Adolescent Health*, 55, 93–99. doi:10.1016/j.jadohealth.2013.11.008
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, 34, 315–346. doi:10.1207/S15327906MBR3403_2
- Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology*, 31, 41–53. doi:0091-0627/03/0200-0041/0
- Dunham, Y., & Olson, K. R. (2016). Beyond discrete categories: Studying multiracial, intersex, and transgender children will strengthen basic developmental science. *Journal of Cognition and Development*, 17, 642–665. doi:10.1080/15248372.2016.1195388
- Eaton, W. O., Von Bargen, D., & Keats, J. G. (1981). Gender understanding and dimensions of preschooler toy choice: Sex stereotype versus activity level. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 13, 203–209. doi:10.1037/h0081181
- Ehrensaft, D. (2011). *Gender born, gender made: Raising healthy gender-nonconforming children*. New York, NY: Workman.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(Serial No. 242), 1–173. doi:10.2307/1166093
- Frey, K. S., & Ruble, D. N. (1992). Gender constancy and the “cost” of sex-typed behavior: A test of the conflict hypothesis. *Developmental Psychology*, 28, 714–721. doi:10.1037/0012-1649.28.4.714
- Goldberg, A. B., & Adriano, J. (2007, April 27). “I’m a girl” : *Understanding transgender children*. Retrieved from <http://abcnews.go.com/2020/story?id=3088298&page=1>
- Green, R. (1976). One-hundred ten feminine and masculine boys: Behavioral contrasts and demographic similarities. *Archives of Sexual Behavior*, 5, 425–446. doi:10.1007/BF01541335
- Halim, M. L., Ruble, D. N., Tamis-LeMonda, C. S., Zosuls, K. M., Lurye, L. E., & Greulich, F. K. (2014). Pink frilly dresses and the avoidance of all things “girly”: Children’s appearance rigidity and cognitive theories of gender development. *Developmental Psychology*, 50, 1091–1101. doi:10.1037/a0034906
- Hidalgo, M. A., Ehrensaft, D., Tishelman, A. C., Clark, L. F., Garofalo, R., Rosenthal, S. M., . . . Olson, J. (2013). The gender affirmative model: What we know and what we aim to learn. *Human Development*, 56, 285–290. doi:10.1159/000355235
- Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology*, 66, 547–574. doi:10.1146/annurev-psych-010814-015025
- Kohlberg, L. (1966). A cognitive-developmental analysis of children’s sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. doi:10.3389/fpsyg.2013.00863
- Levy, G. D., & Haaf, R. A. (1994). Detection of gender-related categories by 10-month-old infants. *Infant Behavior and Development*, 17, 457–459. doi:10.1016/0163-6383(94)90037-X
- Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self*. New York, NY: Plenum Press.
- Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, 67 (Serial No. 269), 1–183.
- Lobel, T. E., & Menashri, J. (1993). Relations of conceptions of gender-role transgressions and gender constancy to gender-typed toy preferences. *Developmental Psychology*, 29, 150. doi:10.1037/0012-1649.29.1.150
- Maccoby, E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
- Marcus, D. E., & Overton, W. F. (1978). The development of cognitive gender constancy and sex role preferences. *Child Development*, 49, 434–44. doi:10.2307/1128708
- Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., & Ruble, D. N. (2016). A dual identity approach for conceptualizing and measuring children’s gender identity. *Child Development*, 88, 167–182. doi:10.1111/cdev.12568
- Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children’s same-sex peer interactions. *Developmental Psychology*, 37, 431–446. doi:10.1037//0012-1649.37.3.431
- Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children’s beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships*, 16, 751–771. doi:10.1177/0265407599166005
- Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*, 60, 870–881. doi:10.1007/s11199-009-9584-x
- Olson, K. R. (2016). Prepubescent transgender children: What we do and do not know. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55, 155.e3–156.e3. doi:10.1016/j.jaac.2015.11.015

- Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science*, 26, 467–474. doi:10.1177/0956797614568156
- Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, 30, 436–442. doi:10.1037/0012-1649.30.3.436
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, 31, 1109–1121. doi:10.1068/p3331
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology*. Vol. 3: *Social, emotional, and personality development* (6th ed., pp. 858–932). New York, NY: Wiley.
- Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, 78, 1121–1136. Retrieved from <http://www.jstor.org/stable/4620693>
- Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, 25, 7–15. doi:10.1080/01650250042000078
- Serbin, L. A., Poulin-Dubois, D., & Eichstedt, J. A. (2002). Infants' response to gender-inconsistent events. *Infancy*, 3, 531–542. doi:10.1207/S15327078IN0304_07
- Shelton, J. N. (2000). A reconceptualization of how we study issues of racial prejudice. *Personality and Social Psychology Review*, 4, 374–390. doi:10.1207/S15327957PSPR0404_6
- Shutts, K., Pemberton, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, 14, 35–62. doi:10.1080/15248372.2011.638686
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, 13, 147–183. doi:10.1006/drev.1993.1007
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. doi:10.1177/0956797611417632
- Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 46, 849–856. doi:10.2307/1128389
- Soh, D. W. (2016, January 4). The transgender battle line: Childhood. *The Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/the-transgender-battle-line-childhood-1451952794>
- Steensma, T. D., Biemond, R., de Boer, F., & Cohen-Kettenis, P. T. (2011). Desisting and persisting gender dysphoria after childhood: A qualitative follow-up study. *Clinical Child Psychology and Psychiatry*, 16, 499–516. doi:10.1177/1359104510378303
- Steensma, T. D., McGuire, J. K., Kreukels, B. P., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 582–590. doi:10.1016/j.jaac.2013.03.016
- Stennes, L. M., Burch, M. M., Sen, M. G., & Bauer, P. J. (2005). A longitudinal study of gendered vocabulary and communicative action in young children. *Developmental Psychology*, 41, 75–88. doi:10.1037/0012-1649.41.1.75
- Szkrybalo, J., & Ruble, D. N. (1999). "God made me a girl": Sex-category constancy judgments and explanations revisited. *Developmental Psychology*, 35, 392–402. doi:10.1037/0012-1649.35.2.392
- Vilain, E., & Bailey, J. M. (2015, May 21). What should you do if you son says he's a girl? *Los Angeles Times*. Retrieved from <http://www.latimes.com/opinion/op-ed/la-oe-vilain-transgender-parents-20150521-story.html>
- Volbert, R. (2000). Sexual knowledge of preschool children. *Journal of Psychology & Human Sexuality*, 12, 5–26. doi:10.1300/J056v12n01_02
- Warin, J. (2000). The attainment of self-consistency through gender in young children. *Sex Roles*, 42, 209–231. doi:10.1023/A:1007039222998
- Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology*, 45, 688–701. doi:10.1037/a0014053
- Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York, NY: Guilford.
- Zucker, K. J., Bradley, S. J., Doering, R. W., & Lozinski, J. A. (1985). Sex-typed behavior in cross-gender-identified children: Stability and change at a one-year follow-up. *Journal of the American Academy of Child Psychiatry*, 24, 710–719. doi:10.1016/S0002-7138(10)60114-8
- Zucker, K. J., Bradley, S. J., Kuksis, M., Pecore, K., Birkenfeld-Adams, A., Doering, R. W., . . . Wild, J. (1999). Gender constancy judgments in children with gender identity disorder: Evidence for a developmental lag. *Archives of Sexual Behavior*, 28, 475–502. doi:10.1023/A:1018713115866
- Zucker, K. J., Bradley, S. J., & Sanikhani, M. (1997). Sex differences in referral rates of children with gender identity disorder: Some hypotheses. *Journal of Abnormal Child Psychology*, 25, 217–227. doi:10.1023/A:1025748032640

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Data S1. Correlation Table and Stereotyping Measure Activities.